APPENDIX I:

ACINETOBACTER BAUMANNNII STRAIN IDENTIFICATION

Acinetobacter baumannii represented the most commonly observed species within samples from this study. This afforded the opportunity to attempt strain characterization in each wound, as strain genotyping could be relevant to clinical outcome in cases where a strain is associated with a particular response. Strain analysis was performed using read mapping by LMAT (Table A1). Briefly, LMAT compares the sequence similarity of each read to all finished and draft assemblies of all bacterial genomes, including *A. baumannii*, and selects the highest scoring match with no conflicting taxonomic matches. A read may be reported to originate from the *A. baumannii* species if the read matches equally well to multiple strains and to no other bacterial species, whereas strain-specific calls indicate greater similarity to a specific strain. Due to low coverage, LMAT was unable to identify reads associated with a distinct strain in five of the samples with *A. baumannii* sequence data (13-2-EBON, 19-2-EA2, 27-2-EA2, 31-1-EA2, and 43-1-WB).

In an attempt to capture *A. baumannii* strain information in samples with lower coverage or fragmented sequence data, an orthogonal approach was undertaken using single nucleotide polymorphisms (SNPs). Analysis was performed using a highly scalable software package (kSNP) to identify SNPs present in sequence data. Briefly, the kSNP approach identifies SNPs from an input sequence by enumerating all possible k-mer oligos within a given target. Candidate SNPs are then identified by examining the central base of each given oligo and determining the incidence of the surrounding sequence within the full set of k-mers. An important advantage of the kSNP platform is that it does not require positional information relative to the whole genome, allowing for analysis of short read metagenomic data. Ability to

process fragmented genome data would be highly useful for analyzing unassembled genomes, particularly with the increasing prevalence of short read sequence information.

Individual wound samples were clustered according to the SNP analysis. Notably, six of the dehisced samples grouped together and were distinct from all other strains, clustering more tightly within these samples than to any single assembled genome (Table A1). Based on LMAT mapping, strain Naval-18 demonstrated the closest association to this cluster. It was not possible, for the available samples in which SNPs were detected, to identify distinct SNPs shared by healed samples that differed from all dehisced samples.

Where coverage was higher, LMAT reported four *A. baumannii* strains that corresponded with distinct SNP clusters. Where sequence data were available, strains associated with healed wounds by LMAT (AB058, TG2026) were distinct from those associated with failed wounds. The failed wounds were grouped into three categories (Naval-18, 6013150, unannotated), and each were associated with different kSNP clusters. It is important to note that the LMAT strain identity assignment did not always agree with the corresponding kSNP assignment, and that not all samples annotated by LMAT were mapped by kSNP (and vice versa). This may be a consequence of differing levels of completeness among the reference draft genomes. Enhancement of coverage could further facilitate the identification of SNPs and strain identities demonstrating associations between healed and failed wounds, which will be the subject of future studies.

Table A1. *Acinetobacter baumannii* strains with closest genetic distance to wound samples.

A. baumannii sequence data were used for genotyping of each wound sample through sequence mapping and SNP analysis. SNP counts were determined using kSNP software. Sequence read mapping was performed using LMAT. For samples in which strains were identified by kSNP but not LMAT, non-strain-specific *A. baumannii* coverage is indicated.

		kSNP analysis		LMAT sequence mapping analysis		
Sample	**Outcome**	**# SNPs (k=19)**	**Closest strain**	**Number of strain-specific reads**	**Closest strain**	**Fraction of genome covered by total reads**
11-1-EBON	Healed	282922	ANC_4097	18,210	AB058	1
34-1-EBON	Healed	0	N/A	4	AB058	0.0008
44-1-EBON	Healed	285432	IS-123	6	TG2026	1
13-2-EBON	Dehisced	101	AB4857	N/A	None	0.0019
16-1-EBON	Dehisced	1041	Wound cluster	13	Naval-18	0.0274
16-2-EBON	Dehisced	268716	Wound cluster	4,267	Naval-18	0.6178
16-2-WA	Dehisced	302	6013113	7	6013150	0.0737
26-1-EA2	Dehisced	468	Wound cluster	7	Naval-18	0.0089
26-1-EB2	Dehisced	1245	Wound cluster	19	Naval-18	0.0516
26-1-EBON	Dehisced	274540	Wound cluster	6,057	Naval-18	1
26-1-ECON	Dehisced	231404	Wound cluster	2,668	Naval-18	0.5489
27-2-EA2	Dehisced	43	Ab33333	N/A	None	0.0009
27-2-EBON	Dehisced	295037	6013150	26,429	6013150	1